Selection Bias in the NBA Draft

Author: Christopher Mattison Perry

Boston College Electronic Thesis or Dissertation, 2008

SELECTION BIAS IN THE NBA DRAFT

C. Mattison Perry
Boston College
Economics Honors Thesis
May 5, 2008

Advised by Prof. Christopher Maxwell

Introduction

The decision making process in labor markets is an economic topic of considerable import. Employers face a number of options in deciding whom to hire and at what wage they are willing to hire those employees. While these decisions may seem straight-forward from a distance, the specific decision to hire one person over another is often a very subjective affair that may well come down to the preferences, tastes, and perhaps biases of the individual making the decision. Economic theory would indicate that for a given job, the person hired is the one with the greatest net Marginal Revenue Product, but in a labor market full of risk, uncertainty, and misinformation it is not always possible to identify the right candidate. Any insight that can be gained into the factors that shape hiring decisions is valuable information for the study of labor economics. Unfortunately, this is a difficult area to study due to the lack of standardized information about whom employers consider for job positions, which candidates they hire for these jobs, and how their subsequent performance compares to expectations. In most industries it is impossible to control or account for all the variables and factors that can skew the analysis of employee hiring. Fortunately the world of professional sports features a richness of data that allows researchers to overcome many of these challenges.

The advantages of analyzing hiring decisions in professional sports derive from the highly controlled hiring markets as well as the fact that individual performance can be more easily measured than in most industries (since athletes do their job in front of thousands of people and their performance is widely reported through the media). Sports statistics are also carefully monitored and are readily accessible for researchers to analyze. In short, the sports market provides a well controlled environment that is highly conducive to the study of labor economics in general, and hiring decisions in particular. While many of the factors that shape hiring decisions are specific to individual sports, some of the decision making trends echo similar trends in other hiring markets. A review of the literature in this area reveals a variety of hiring themes that are analyzed (For more on labor economics in the sports industry, see Kahn, 2000)

In this paper I analyze the NBA hiring market, specifically the NBA Draft. My objective was to create a model that detects bias in the way different groups of players are drafted. To this end I will present control models that predict the eventual value of players based on where they are selected in the Draft. If a group of players consistently over-performs relative to their expected value, as predicted by the model, then I conclude that the group in question is experiencing negative bias in the Draft. The opposite also holds if a group consistently under-performs relative to their predicted value. In this scenario I conclude that there is a bias in that group's favor. Using this method I detected bias related to high school players, centers, black players, and foreign players. The nature of each of these biases is unique. The bias works in favor of some groups while it works against others. Some groups experience bias throughout the Draft while others are specific to one round, or to certain years in the data[1]:

- High school players are discriminated against in the Draft, especially when they are selected with a lottery pick
- Centers experience the opposite effect in the lottery and are consistently drafted too high in that part of the Draft
- Black players are consistently drafted too low in the first round, and then too high in the second round

[1] Appendix A shows where the players from these groups were drafted from 1995-2003.

- Foreign players were drafted too low from 1995-2001, and then too high in 2002 and 2003

Each of these effects point to unique trends in the NBA labor market.

The rest of this paper will proceed as follows. In the next section I briefly review the literature on discrimination in sports, focusing on the interplay of risk aversion and option value, Draft order bias, and racial bias. Next I explain my method for quantitatively valuing players and comparing player performance across years. From there I present my control models which give an expected value for every player in the Draft as a function of which pick they are selected with. After an explanation of some of the issues I encountered and how I controlled for different data problems, I present my results for each of the four groups. These results lead into the conclusion.

Literature Review

There is a great deal of literature that focuses on the economics of discrimination. One of the most important, and earliest, works in this area is Gary Becker's 1957 dissertation, "The Economics of Discrimination". The literature on discrimination in sports is nicely summarized in Lawrence Kahn's 1991 survey, "Discrimination in Professional Sports: A Survey of the Literature" as well as his 2000 paper mentioned earlier. His work has focused primarily on discrimination in pay, hiring, and promotion as well as the institution of "stacking". Though some authors have looked at discrimination in the player Draft, that dimension of discrimination has received far less attention than the other forms. A review of the literature in this area reveals a variety of hiring themes that are analyzed.

1. Risk Aversion, Option Value, and Draft Order Bias

A major issue in any market involving uncertainty is the interplay of risk aversion and option value. How one chooses to manage risk will have an effect on any hiring decision. This effect is particularly pronounced in the sports hiring market where the level of uncertainty is fairly high.

The effect of risk aversion and option value on hiring practices has been explored in detail in the context of the NFL Draft by Hendricks, DeBrock & Koenker (2001). In that paper the authors compare two groups of potential NFL players. Those players who played for Division 1 schools are considered the low risk group while all other players are placed in the higher risk group. The study found that teams were risk averse and demonstrated a preference for drafting low risk players (who had a high likelihood of being successful professionals) in the early rounds. The result was that on average when controlling for Draft position, non-Division 1 players had careers that were more successful (since the drafting teams paid a premium for low risk Division I players). This risk aversion caused non-Division I players to be drafted later, which led to them outperforming their Draft position as a group. In the later rounds, this trend shifted since fewer drafted players were likely to become successful pros. In these rounds more value was placed on a player's option value. Division 1 players drafted in these rounds had better careers, on average, than their non-Division 1 counterparts since teams would rather "gamble" on a high variance player at that point in the Draft.

A study by Groothius, Hill & Perri in 2005 found that option value plays a role in the NBA Draft as well. With the new Collective Bargaining Agreement (CBA) in 1995, the NBA instituted the rookie pay scale for first round picks that locked players into set three-year contracts. These contracts often paid less than a player's MRP and created a window for players to be evaluated before signing a large contract. Given relatively inexpensive labor and this extended period over which to evaluate players, the authors found that teams had an added incentive to pursue riskier players. My analysis will further explore the role of risk and option value in the NBA Draft.

2. Racial Bias in Sports Hiring

Another hiring trend that can be analyzed through the lens of sports is racial bias in hiring markets. Racism has played a significant role in American history and continues to be a major issue of concern. Economists have looked at this issue in the context of the NBA and some have found significant results. In his paper, "Racial Differences in Professional Basketball Players' Compensation", Lawrence Kahn analyzed NBA salaries for the 1985-86 season and reached an interesting conclusion. He found that though black and white players received equal compensation on average, this was no longer true once he controlled for performance. On average black players performed at a higher level than whites, and when his salary model controlled for this fact he found that black players were underpaid by about 20% compared to their white counterparts. He theorized that this bias was able to go largely unnoticed since the average compensation was the same for both groups. Had black and white players in the NBA had the same average ability then it would have been much more glaring that white players were making an average of 20% more (Kahn & Sherer, 1988). This evidence points to some kind of discrimination, but it is important to identify the specific form of discrimination at work.

As Kahn points out, if the discrimination stems from preferences of owners, coaches, or other players then the effect would be to drive black players towards certain teams that did not discriminate. We can realistically assume that these teams exist, given that the majority of players in the league are black and there are also many black coaches. The discriminating teams would be willing to pay a premium for white talent while the non-discriminating teams would get higher quality labor at a cheaper price. The result of this is a system that forces the discriminating teams to either change their ways or continually underachieve relative to their more inclusive counterparts. This scenario does not appear to fairly describe the NBA. Kahn proposes instead that the driving factor is consumer discrimination. In defining consumer discrimination, Gary Becker stresses the difference between two individuals being "perfect substitutes in production" versus being "perfect substitutes in marketable production" (Becker, 1971). While a black player and a white player may have identical skills, if a largely white fan base prefers watching the white player then his MRP will be higher based on his greater marketable production. Unlike the situation where teams are the discriminating agents, this scenario is not self-correcting. As long as this sort of fan bias exists, it will be profitable to pay a premium to have white players on your team. According to Kahn, this was the source of the 20% pay disparity.

One peculiarity in Kahn's results is that when he analyzed the NBA Draft he found that black players were selected higher on average than their white counterparts, and when he controlled for ability he found no racial Draft order bias. While this supports the idea that the bias comes from the fans and not the team, it points to an inconsistency in the way teams were being run at that time. In drafting players, teams seemed to value talent above all else, and in doing so they drafted in a way that may be considered more ethical by some. However if we believe customer discrimination is significant in the NBA, then we can also say that the teams' drafting behavior was not profit maximizing. On the other hand, when signing free agents, teams sought to maximize profits by overpaying marketable white players. This approach does not appear to be a rational equilibrium strategy, and I suspect that analysis of more recent data would reflect different behavior. Assuming that teams seek to maximize profits, over time it stands to reason that there has either been a decline in consumer discrimination or teams have begun to favor white players in the Draft over equally talented black players. If there really is a premium on white players then it should be reflected in the Draft as well as in free agency.

More recent evidence indicates that there may still be consumer bias among NBA fans, but not nearly as much as was found for 1986. A study based on the 2001-2002 season found shortfalls in salary, total compensation, and contract duration for marginal black players versus their white counterparts. The results did not, however, show a significant effect among free agents or players under their rookie contracts (Kahn & Shah, 2005).

Looking beyond salaries, there is more evidence for consumer bias. A study looking at Nielsen ratings for televised NBA games found that, other things equal, more fans tune in to games when there are a greater number of white players on the competing teams (Kanazawa & Funk, 2001). Accordingly, it should be more profitable to field a white team over a black team of equal talent. This creates a market environment that encourages pay discrimination.

If the market for NBA talent is partially driven by bias, one would expect this discrimination to be reflected in the hiring process. Though the evidence for racial bias in the NBA today is not all that strong, it is certainly possible that race plays a role in shaping hiring decisions.

The NBA Draft

The NBA Draft has been a part of the NBA since the league's formation in 1949. An annual event held in June, its purpose is to distribute the best available amateur talent to each of the teams in the league. As the league has evolved over time, the rules of the Draft have changed fairly significantly, but it has existed in more or less the same form since 1995. The Draft provides a unique opportunity to analyze the hiring decisions of each team, and how a player's performance compares to his anticipated value on Draft day. The pool of available talent consists of college seniors who have finished their college basketball careers, underclassmen and high school seniors[2] who forego their college eligibility and declare themselves eligible for the Draft, and international players who choose to make themselves Draft eligible. Today, the Draft consists of two rounds with each team receiving one pick per round. With just one exception, the order of the Draft is determined by each team's record during the previous season. The teams with the lowest win totals select first in an attempt to maintain competitive balance within the league. The order of both rounds is set by this reverse order of finish with the exception of the first three picks of the first round. These picks are distributed by a lottery that includes all of the teams that fail to make the playoffs in the previous season.[3] This lottery is heavily weighted so the weakest teams of the group have a better chance of receiving these high picks. However, since all the teams that miss the playoffs have some chance of getting these lottery picks they are often referred to as "lottery" teams, a term that I will be using frequently in this paper. Once the top three picks are set, the rest of the round continues based on the previous year's win-loss record.

When a player is drafted, that team has exclusive rights to sign the player to a contract. Players selected in the first round receive a guaranteed three year contract[4] that pays them a pre-

[2] Due to a recent rule change, high school seniors are no longer eligible to be drafted; however this rule was not in place during the time of my study. High school players play an important role in my analysis.
[3] There are currently 14 such teams in the lottery since 16 of the 30 NBA teams make the playoffs. However, for the time of my study there were only 29 NBA teams so the lottery consisted of 13 teams.
[4] Starting with the 1998 Draft, teams also had an option to keep players on their rookie contract for a fourth year. This rule change will resurface as something I control for when running my regressions.

determined salary based on their Draft position.[5] The higher a player is drafted, the more money he makes; but there is no negotiation involved due to this pre-set rookie salary scale. In most cases, these rookie contracts pay less than a player's true market value. Once these contracts expire, players become "free agents" and are free to sign with the highest bidder. Only then will they get paid in a manner that reflects their true worth. Second round picks do not enjoy the same job security as players taken in the first round. These players receive no guaranteed contracts and have to earn their roster spots before signing a deal. Many of these players never make it in the NBA, and those that do typically sign one or two year contracts for the league minimum. If they are successful, they are re-signed at a price that accurately reflects their market value.

Data

In order to run an NBA Draft-order bias model, I had to collect a significant amount of data. Fortunately sports data is widely available. My analysis focuses on NBA players drafted from 1995-2003. The required salary data is available on Patricia Bender's highly regarded basketball website as well as the USA Today online salary database. Ms. Bender's site also gives a history of NBA Draft picks and basic information about the players selected. This data includes where the player was selected, where they played basketball previously (high school, college, international, etc.), and their position on the court.

Measuring Player Value

The first step in creating a model of Draft-order bias is to identify some quantitative measure of player value that allows us to compare the relative value of different players. Luckily, the structure of the Draft creates an environment that is conducive to just such a comparison.

[5] Appendix B features a sample rookie contract scale from the 2003 Draft.

Since the rookie contracts of players drafted in the first round are standardized, they do not provide a useful way to compare players. The only thing these contracts reflect is the player's Draft position. These contracts do, however, create an evaluation period in which each player can be observed before they become free agents and receive a market-value contract. Over this time, all of the teams in the league have an opportunity to gauge each player's worth. Therefore, when players eventually re-sign, we can well assume that their salary is reflective of their true market value. The other benefit of the rookie contract is that all of the first round picks begin their second contracts at the same time. Players selected in the second round never have scaled contracts and are already receiving their market value by the time the first round picks re-sign, assuming they are still in the league. For my analysis, I use the salary that each player received in the year after their Draft class completed their rookie contracts as the quantitative measure of that player's value.

There are other approaches that I considered as a measure of value, but each had its own pitfalls. One completely different approach would be to use performance statistics from games as a way of rating players, and to compare value in this way. This would be a complicated process and the values it produced could be distorted by what kind of role a player is asked to play. Further, any intangible things that a player does on the floor are not picked up by statistics. Ultimately, it is always reasonable to assume that employees in competitive labor markets are paid what they are worth and any approach that wasn't based on salary would be unnecessarily indirect.

I also considered using career earnings to measure worth. The flaw here was that a player whose career was cut short by injury would have career earnings that were not reflective of his talent level. Since the purpose of this analysis is to compare pre-Draft expectations to post-Draft performance, one cannot expect teams to foresee injuries late in a player's career when they are drafting that player.

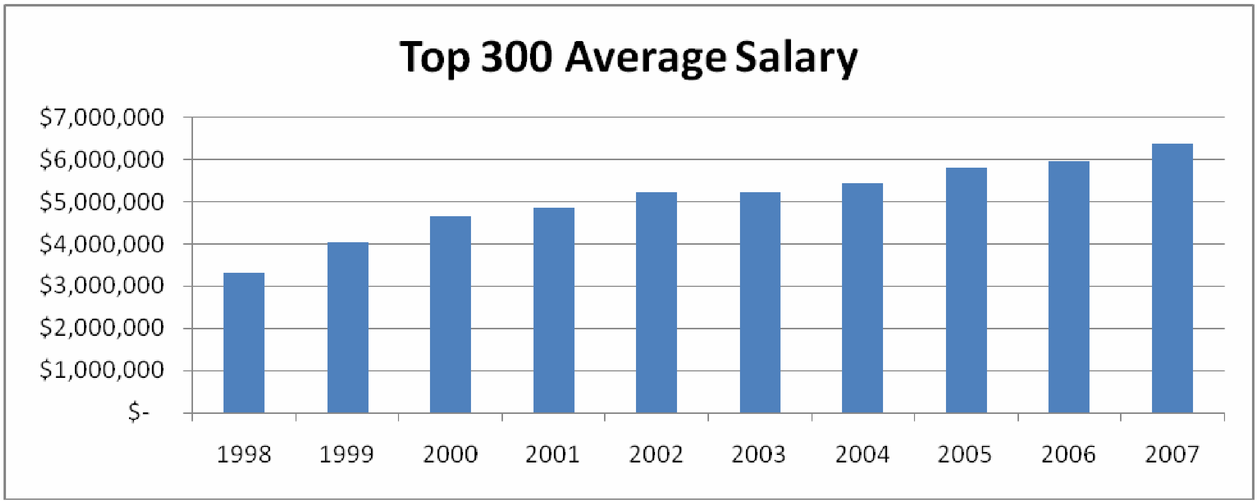
Another possible approach is to employ a player's average career salary, but this would misjudge players who play past their prime for less money. Further, it would reward players who improve later in their careers (improvement that would not be foreseeable to teams on the night of the Draft). These approaches were all reasonable, but ultimately the post-rookie salary provides the best form of comparison between players from a given Draft year.

Standardizing Player Salaries

To proceed with the model, it is not sufficient to simply compare players to others from their own Draft class; we must also be able to compare players across years. As mentioned above, the data I use for this study consists of players drafted from 1995-2003. I do not go back further than 1995, because that is the first year in which there was a set scale for rookie contracts. Prior to 1995, players negotiated their contracts freely with the teams that drafted them. This eliminates the evaluation period that exists in the current format; hence any comparison to years before 1995 would bring a whole series of other variables into play. The last Draft class in my study is 2003, because that is the most recent Draft class that has signed new contracts.

In order to compare all of these players in one model, I standardize their salaries by adjusting for inflation. To do this, I took an average of the salaries of the top 300 players in the league for every year, starting in 1998[6].

[6] This is the year in which the Draft class of 1995 signed their post-rookie contracts.



	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Avg. Salary ($)	3,329,798	4,061,828	4,652,584	4,878,173	5,237,556	5,226,058	5,452,104	5,815,815	5,954,811	6,404,399
% of 1998	100%	122%	140%	147%	157%	157%	164%	175%	179%	192%

Using the percent change in average salary from year to year, I deflated the contracts signed in every year after 1998 using the index value corresponding to that year. With standardized salaries for the period in question, it is now possible to construct a model that predicts performance based on Draft position for the Draft classes from 1995-2003.

Draft Position Effects

The control model for this analysis should predict the eventual salary of a player selected anywhere in the Draft based on when they were drafted. There will certainly be players who deviate from this expected performance, given the amount of uncertainty that is inherent in the Draft, but for any group of players we would expect to see that deviation distributed evenly. That is to say, for a given group of players, there should be as many individuals who over-perform relative to where they are selected as there are players that under-perform. Any group that consistently falls on one side of this would appear to be experiencing some sort of bias.

There are a number of ways to control for Draft position effects. Ultimately, I considered three Draft position control models. Each Draft position control model takes a unique shape, but ultimately they all generate similar results.

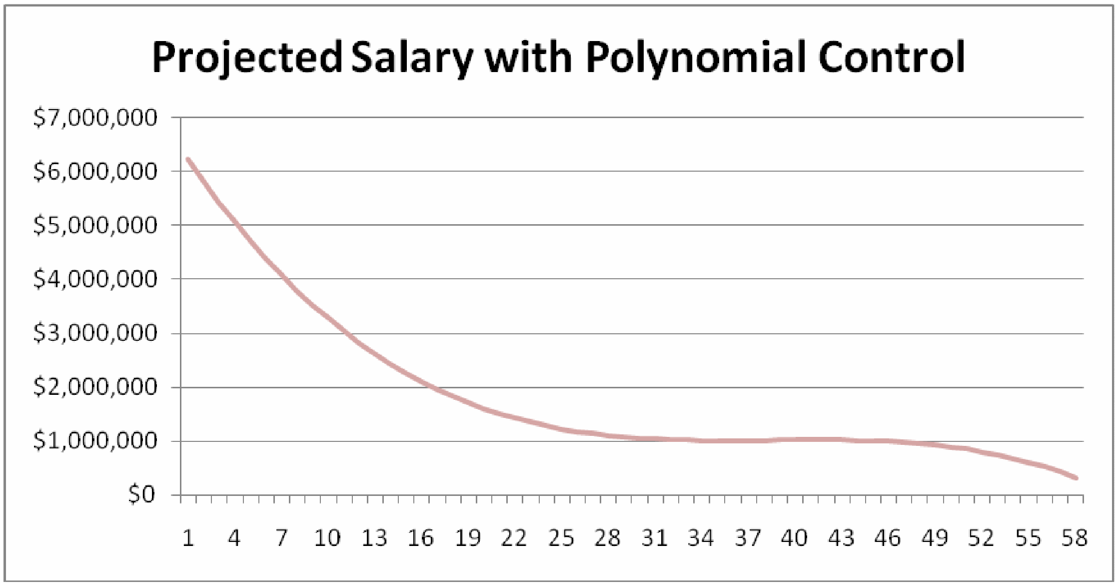
1. <u>Polynomial Control</u>

The first control model I employed was a simple third-order polynomial that fits a smooth curve to the data. The variables I used were the position in which a player was picked (Pick), that number squared (Pick^2), and that same number cubed (Pick^3). I also included dummy variables for the Draft classes from 1997, 2000, and 2002, as those Drafts were particularly weak and the model needs to control for this.[7] Using this polynomial model, I attained the following results[8]:

Polynomial Control Model	
Pick	-$412,418 (8.96)
Pick^2	$9,941 (5.58)
Pick^3	$82 (4.16)
1997	-$515,607 (2.20)
2000	-$574,336 (2.46)
2002	-$400,355 (1.67)
Constant	$6,797,664 (20.82)
Chi^2	334.53

[7] The coefficient for each of these dummy variables is negative and statistically significant. These dummies are intuitive since a weak talent pool will have a negative effect on every pick, bringing each value down.
[8] All interval regression results have 514 observations, 255 uncensored observations, 29 right-censored observations, and 230 interval observations. Z-values are shown in parentheses.

When placed on a graph, with Draft position on the X-axis and adjusted salary on the Y-axis, the Polynomial control model looks like this:[9]



This smooth path cuts through what is a jagged set of results. One issue with this model is that it briefly curves upward in the early second round. This movement violates an obvious and basic assumption that it is always better to pick earlier in the Draft. Overall, the movement is minor and does not take away too much from the model. The next control model takes a different approach.

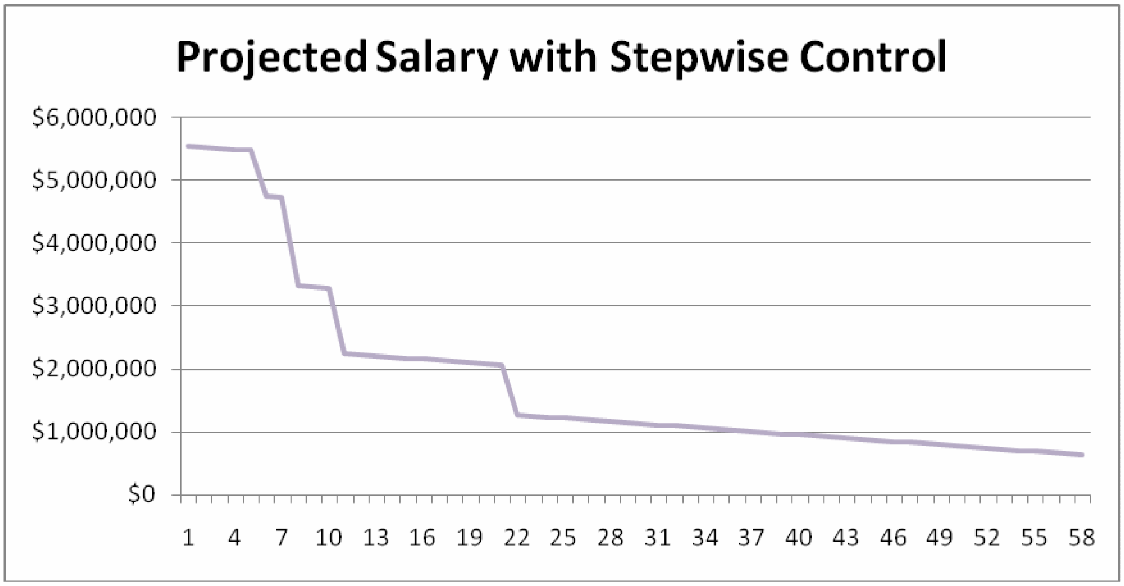
2. <u>Stepwise Control Model</u>

My second control model is stepwise linear with steep drops that reflect those in the actual data. This model assumes that there are a series of steps, where the talent is fairly similar, and then steeper drops down to the next level. The levels selected capture the significant drop-offs in the average salary of players selected at each position. This model employs a set of dummy variables with dummies for a top 5 pick, picks 6-7, picks 8-10, and picks 11-21. In

[9] The graph is applicable for any year besides 1997, 2000, and 2002. For these years you would subtract the appropriate coefficient and shift the graph down accordingly.

addition to these dummies I included a ‘pick’ variable, which corresponds to where a player was selected, in order to pick up linear trends in the relationship between Draft pick and value. As in the Polynomial model, I added dummy variables for each of the three weak Draft years. This model produced the following results:

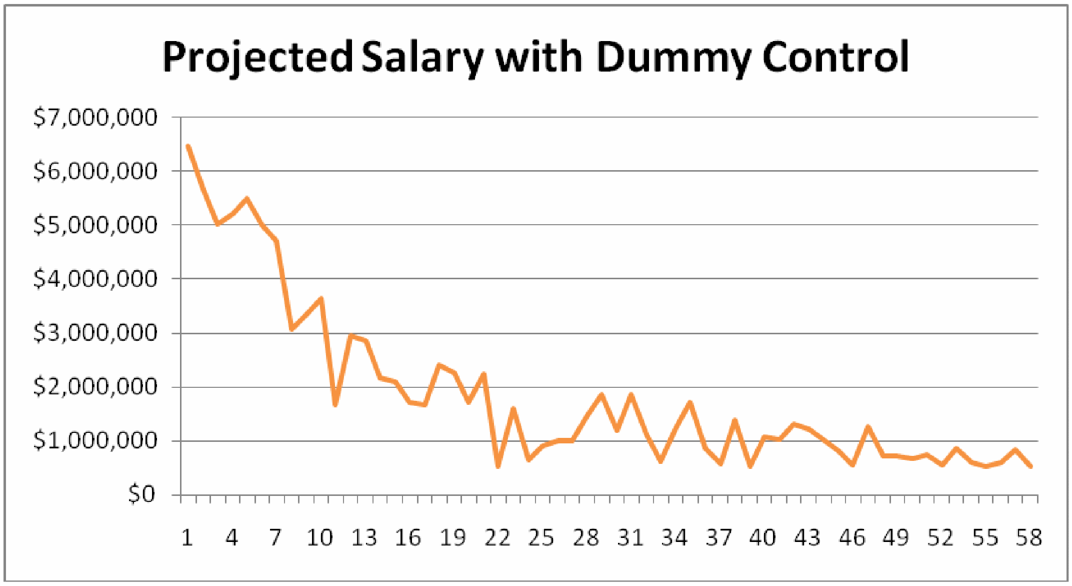
Stepwise Control Model	
Top 5 Pick	$3,934,576 (9.67)
Picks 6-7	$2,785,367 (5.60)
Picks 8-10	$1,910,789 (4.62)
Picks 11-21	$752,407 (2.78)
Pick	-$27,991 (3.41)
1997	-$519,041 (2.25)
2000	-$575,495 (2.51)
2002	-$405,821 (1.72)
Constant	$2,068,453 (6.00)
Chi^2	351.29



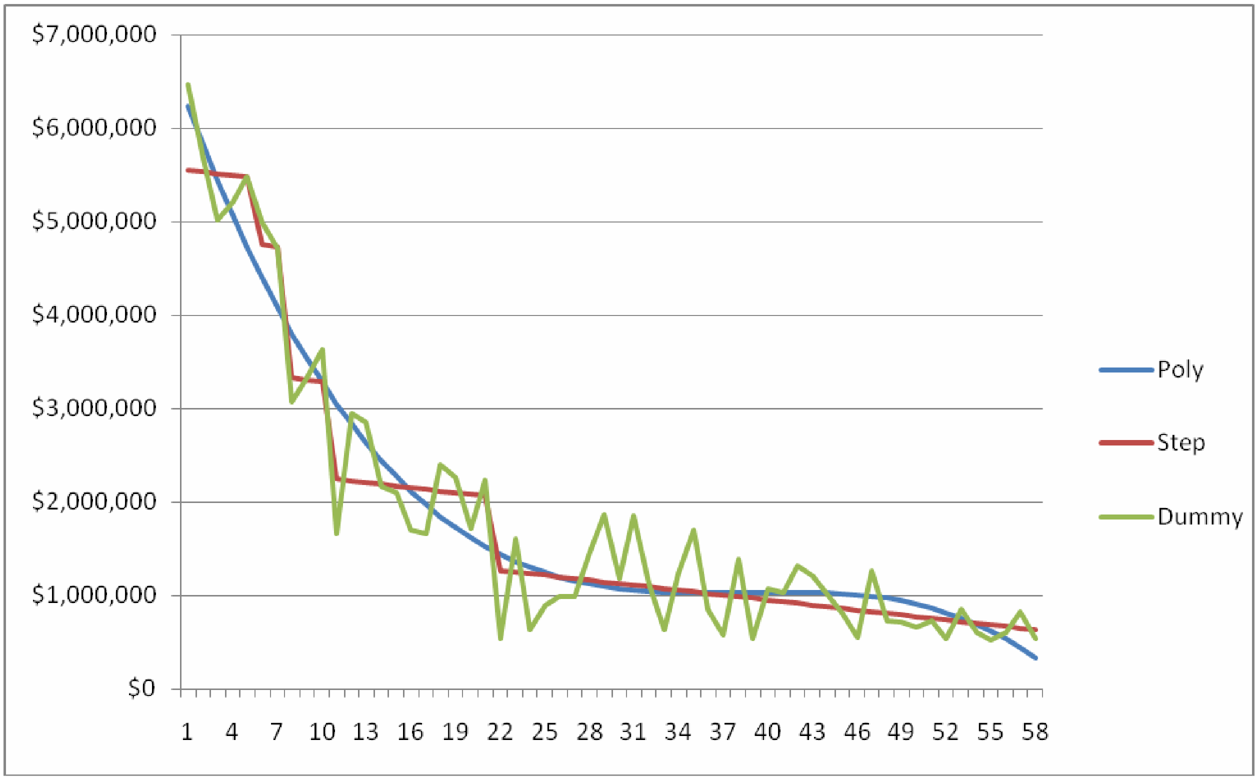
This graph presents a somewhat different profile than the Polynomial model, though the values are similar. The Stepwise model has the added advantage of being a strictly decreasing model. The final control model reveals the chaos that the prior two models attempt to fit.

3. Dummy Control Model

In order to create the most rigorous control possible, this final model simply uses a dummy variable for each position in the Draft (excepting one, so as to avoid co-linearity). The graph of the Dummy control model therefore just reflects the average salary achieved by the players who were drafted in each Draft position. The graph of these results is much less orderly:



The graph reflects the great deal of error and uncertainty that are a staple of the drafting process. While this control model provides the strongest test for the independent variables that will be included in the model, with a chi^2 value of 376.54, it is also the most flawed control model. A fundamental assumption of this analysis is that it is always better to have an earlier pick in the Draft. This much is obvious, since the pool of players is larger the earlier you pick. However, this dummy control model violates this assumption in places where the projected value of one pick is lower than the projected value of the next. One example of this is the difference between the 11th pick, which spikes down and the 12th pick, which has a higher value. Nobody would argue that the 12th pick is preferable to the 11th, yet that is what this model implicitly assumes. This anomaly, which can be observed with other picks as well, is clearly a result of random chance and the limited sample size of just nine years. In contrast, the Stepwise model fits the data in a strictly decreasing fashion and the Polynomial model adheres to this rule except for the brief stretch where it increases slightly in the second round.



It is evident from this overlay that these three models track each other fairly closely. As I will show in the next section, the regression results are similar regardless of which Draft position control model is used.

The Base Combination Model

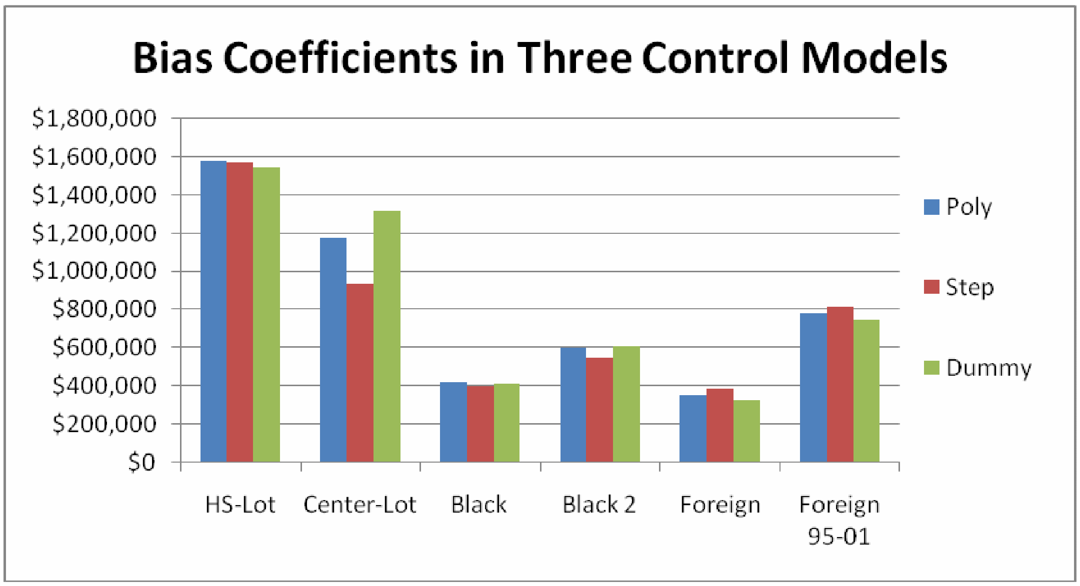
The results I will be reporting below are based on a base combination model that includes the control variables just discussed, as well as six additional dummy variables that I found to pick up on statistically significant trends in the way that players are drafted. They are:

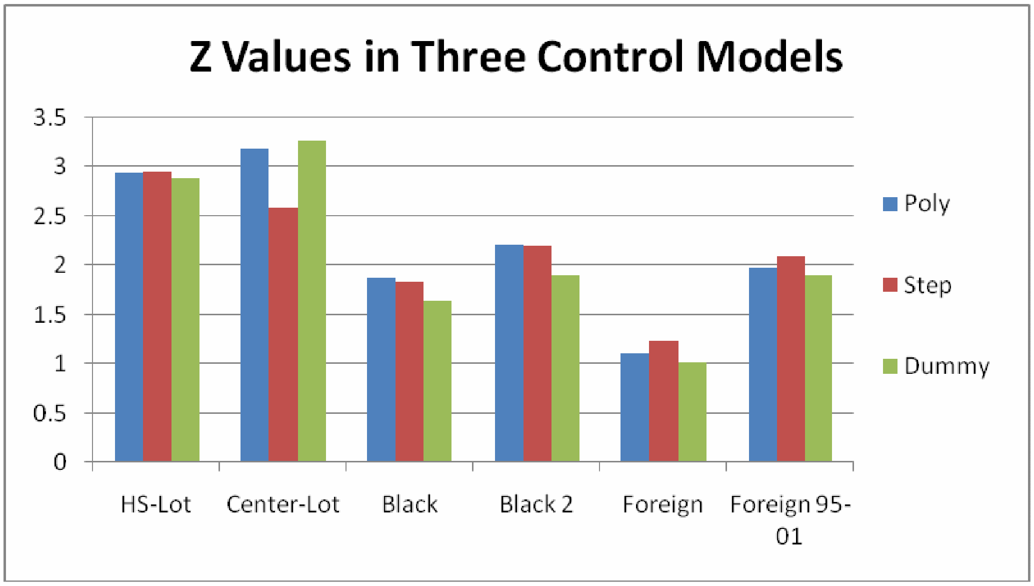
- High school lottery picks
- Centers picked in the lottery
- Black players
- Black players taken in the second round

- Foreign players
- Foreign players drafted from 1995-2001

The base combination model includes all of these variables simultaneously since some of the players fall into multiple categories and just including one variable at a time will not successfully isolate the effects.

Using this base combination model I compared the results from the three control models and found that the coefficients and z-values were generally very similar.



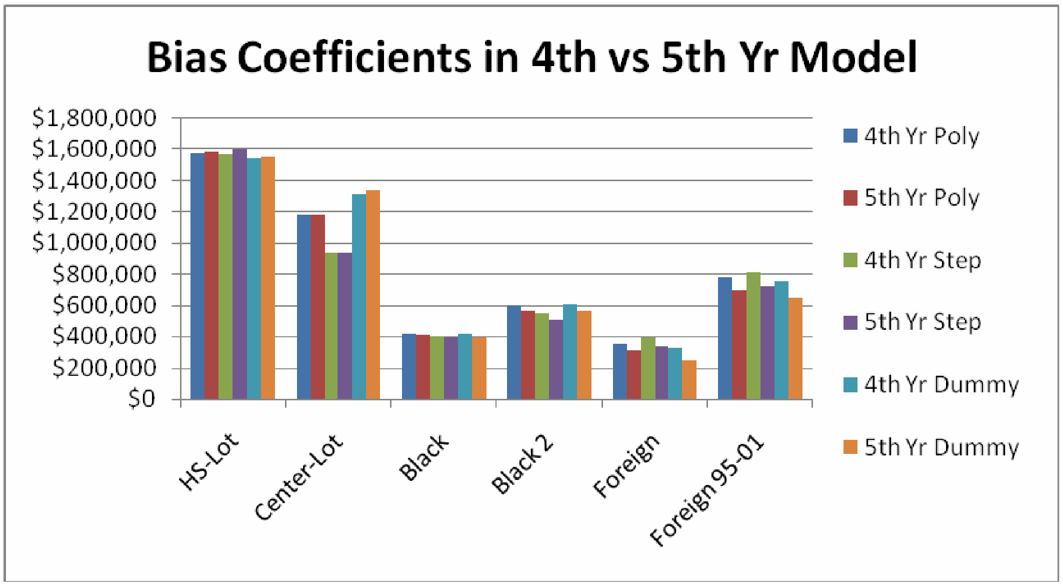


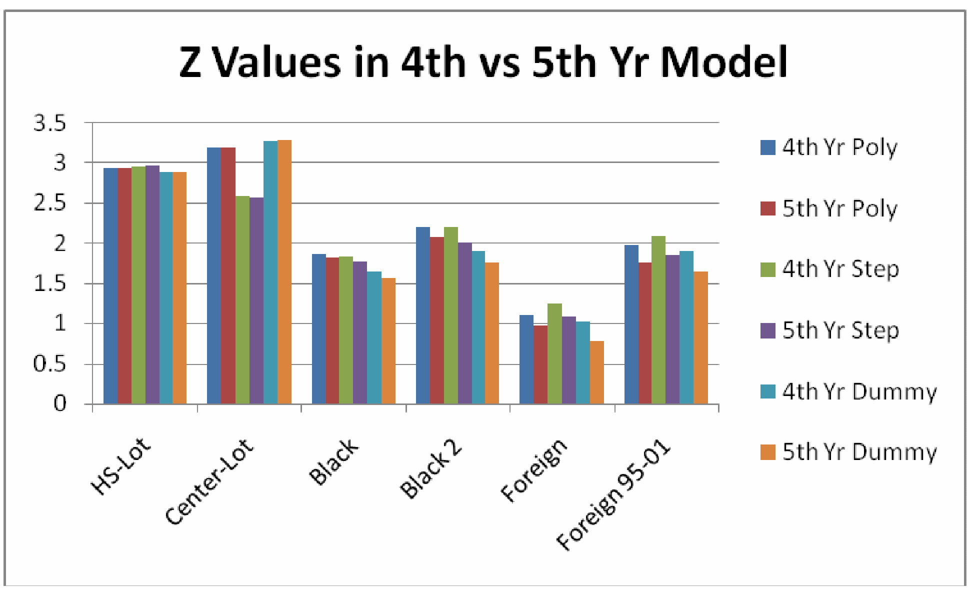
Note that the over-fitted Dummy model produces results very similar to those of the other two. The only notable disparity to be found is in the center-lottery effect. This is likely a result of how the Stepwise model flattens out the projections for top five picks. In this regard the other Draft position control models may be more on target, but either way the z-value for centers in the lottery is statistically significant. The majority of my analysis will focus on the results from the Stepwise control model

<u>Data Issue I: Length of Rookie Contracts</u>

Earlier, I mentioned a slight adjustment to the rookie contract rules between the 1997 and 1998 Drafts. Rookie contracts had been guaranteed for three years, but starting in 1998 they were guaranteed for three years with a team option to retain the player for a fourth year. In most cases, this option was utilized, since the rookie pay scale tends to pay players less than their market value. This creates an inconsistency in the analysis where, from 1995-1997, I evaluate players based on their salary in their fourth year, while from 1998-2003, my evaluation comes from the players' fifth year salary. In general, players make more money each year that they are in the league, so it was possible that this difference might skew the results.

To test for this possibility, I first ran models for each of my controls with relevant variables included. I then re-ran the same models, using adjusted fifth year salary for the players from 1995-1997, instead of the adjusted fourth year salary, as I had been doing:





As you can see from these results, there is a negligible difference in both coefficients and z-values when the model uses fifth year salaries for the 1995-1997 draftees. The rookie contract length proves to be a non-issue in skewing the results.

Data Issue II: Salary Censoring

A much more pressing issue than rookie contract length was the problem of censored salary data. This censoring happens in two different ways, at opposite ends of the pay spectrum.

1. Maximum Contracts

The first issue is that the NBA sets maximum contracts for individual players. The table below presents these values by year and years of NBA experience:[10]

Years in NBA	98-99	99-00	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08
0	$9,000,000	$9,000,000	$9,658,000	$10,625,000	$10,067,750	$10,960,000	$10,968,000	$12,000,000	$12,455,000	$13,041,250
1	$9,000,000	$9,000,000	$9,658,000	$10,625,000	$10,067,750	$10,960,000	$10,968,000	$12,000,000	$12,455,000	$13,041,250
2	$9,000,000	$9,000,000	$9,658,000	$10,625,000	$10,067,750	$10,960,000	$10,968,000	$12,000,000	$12,455,000	$13,041,250
3	$9,000,000	$9,000,000	$9,658,000	$10,625,000	$10,067,750	$10,960,000	$10,968,000	$12,000,000	$12,455,000	$13,041,250
4	$9,000,000	$9,000,000	$9,658,000	$10,625,000	$10,067,750	$10,960,000	$10,968,000	$12,000,000	$12,455,000	$13,041,250
5	$9,000,000	$9,000,000	$9,658,000	$10,625,000	$10,067,750	$10,960,000	$10,968,000	$12,000,000	$12,455,000	$13,041,250
6	$9,000,000	$9,000,000	$9,658,000	$10,625,000	$10,067,750	$10,960,000	$10,968,000	$12,000,000	$12,455,000	$13,041,250
7	$11,000,000	$11,000,000	$11,589,000	$12,750,000	$12,081,300	$13,152,000	$13,161,000	$14,400,000	$14,946,000	$15,649,500
8	$11,000,000	$11,000,000	$11,589,000	$12,750,000	$12,081,300	$13,152,000	$13,161,000	$14,400,000	$14,946,000	$15,649,500
9	$11,000,000	$11,000,000	$11,589,000	$12,750,000	$12,081,300	$13,152,000	$13,161,000	$14,400,000	$14,946,000	$15,649,500
10+	$14,000,000	$14,000,000	$14,000,000	$14,875,000	$14,094,850	$15,344,000	$15,355,000	$16,800,000	$17,437,000	$18,257,750

Players earning the maximum are the NBA's elite. As a result there were only 29 players who fell into this category, out of the 514[11] total players in the dataset. The statistical issue that arises

[10] Table is from cbafaq.com
[11] 514 players represents every player drafted from 1995-2003 except for five that I omitted from the data for various reasons:
-Dermarr Johnson ('00, #6) was in a near fatal car crash that derailed his career while under his rookie contract.
-Jay Williams ('02, #2) was in a near fatal motorcycle accident after his rookie year that ended his career.
-Dajuan Wagner ('02, #6) suffered a debilitating colon condition after his rookie year that ended his career.

is that since there is a maximum, these player's salaries do not reflect their true market value. When a player earns the maximum we only know that he is worth some amount that is greater than or equal to what he is paid. A simple ordinary least squares regression (OLS) will value that player at the statutory maximum, which likely undervalues his actual worth. This problem represents the upper half of the censoring issue.

2. <u>NBA Dropouts</u>

The lower half of the censoring problem occurs with players who do not make it beyond the rookie contract window and who are therefore earning zero NBA dollars during the year in which I evaluate them. Since there is a league minimum salary, we know that these players are worth something between zero dollars and the league minimum, but we have no way of being any more specific. Furthermore, there is the issue of players at the league minimum, who might have a true worth lower than that value, but their salary is inflated by the artificial wage floor. The league minimums, by year and years of NBA experience, are in the table below:[12]

Years in NBA	98-99	99-00	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08
0	$287,500	$301,875	$316,969	$332,817	$349,458	$366,931	$385,277	$398,762	$412,718	$427,163
1	$350,000	$385,000	$423,500	$465,850	$512,435	$563,679	$620,046	$641,748	$664,209	$687,456
2	$425,000	$460,000	$498,500	$540,850	$587,435	$638,679	$695,046	$719,373	$744,551	$770,610
3	$450,000	$485,000	$523,500	$565,850	$612,435	$663,679	$720,046	$745,248	$771,331	$798,328
4	$475,000	$510,000	$548,500	$590,850	$637,435	$688,679	$745,046	$771,123	$798,112	$826,046
5	$537,500	$572,500	$611,000	$653,350	$699,935	$751,179	$807,546	$835,810	$865,063	$895,341
6	$600,000	$635,000	$673,500	$715,850	$762,435	$813,679	$870,046	$900,498	$932,015	$964,636
7	$662,500	$697,500	$736,000	$778,350	$824,935	$876,179	$932,546	$965,185	$998,967	$1,033,930
8	$725,000	$760,000	$798,500	$840,850	$887,435	$938,679	$995,046	$1,029,873	$1,065,918	$1,103,225
9	$850,000	$885,000	$923,500	$965,850	$1,000,000	$1,000,000	$1,000,000	$1,035,000	$1,071,225	$1,108,718
10+	$1,000,000	$1,000,000	$1,000,000	$1,000,000	$1,030,000	$1,070,000	$1,100,000	$1,138,500	$1,178,348	$1,219,590

-Nenad Krstic ('02, #24) did not join his team until a few years after being drafted and is still on his rookie contract.
-Carlos Delfino ('03, #25) similar situation to Krstic, also still on his rookie contract.

[12] Table is from cbafaq.com

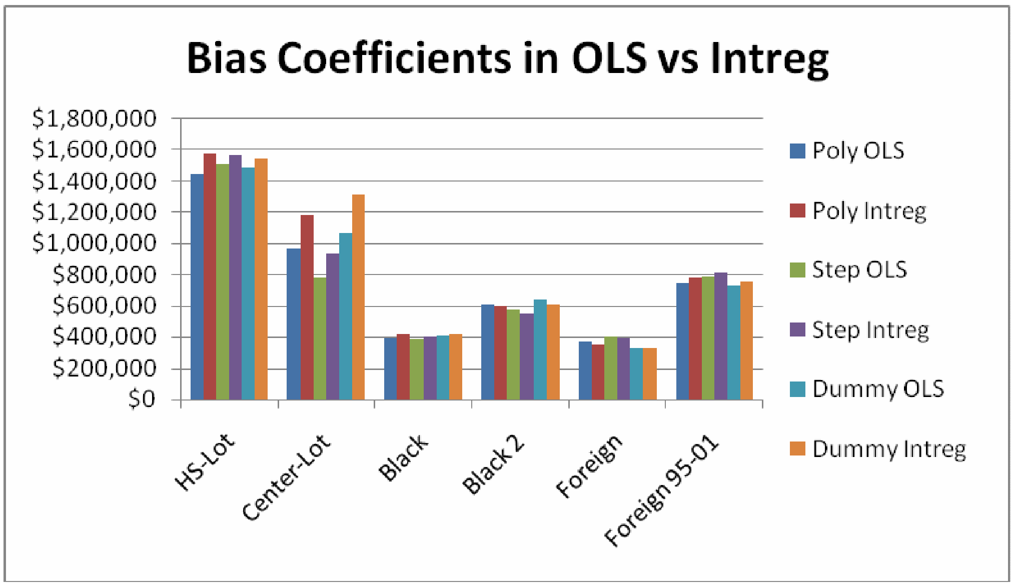
This problem is compounded by the large number of players to whom it applies. Of the 514 players in the data, 230 (about 45%) were no longer in the league by the end of the rookie contract window. In an OLS model, these players are all treated as having an NBA market value of zero dollars.

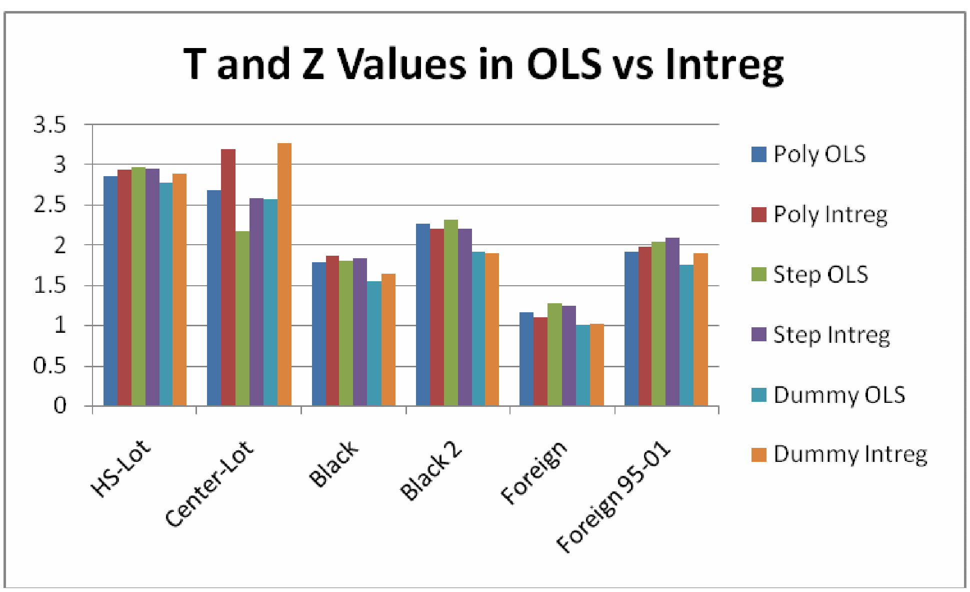
Interval Regression Model

The answer to each of these censoring issues is to run an interval regression. The interval regression is effectively able to process both points and intervals as the dependent variable, using maximum likelihood estimation to fit the model to these values.[13] In the interval regression approach, any player not subject to the above salary constraints is simply entered as a point value, as they would be in an OLS regression. Players earning the maximum have a lower bound that is their adjusted salary, and an indefinite upper bound. In this way, the model allows for their market worth to exceed their salary level. At the lower end of the salary spectrum, I first made the assumption that players earning the minimum were actually worth that amount. I felt safe in this assumption because of the large number of players who are drafted and do not make it in the NBA. With such a large pool of players fighting for limited spots at the bottom of the league's pay chart, one would expect that the players who are signed are actually worth that amount, or at least very close to it. By making this assumption, I was able to treat players earning the minimum as point values in the maximum likelihood estimation and separate them from players who are not in the league. In the interval regression model, NBA dropouts are treated as having an NBA market value in the interval between zero and the league salary minimum for that player. These intervals differ from player to player. Since the NBA's minimum salary is dependent on how many years a player has been in the league, players who were in the league a

[13] See Appendix C for a more detailed explanation.

few years, but did not play past their rookie contract window, are subject to a higher minimum salary than players who never signed with a team. Accordingly, the interval regression model distinguishes between these players by creating a higher ceiling on the interval for a player who played in the league for longer. This modeling flexibility significantly improves the way the model represents actual player value. As a representation of reality, the interval regression seems far superior to the OLS. However, when we compare results, we find very similar coefficients, as well as t and z-values, in the OLS and interval regression models.





Where there are small differences, the interval regression model tends to generate larger coefficients with greater statistical significance. I believe this is attributable to the underlying reality of the trends that this model picks up, and the fact that the interval regression model is simply more statistically appropriate than the OLS approach.

Specification of the Draft-Order Bias Control Model

In order to proceed with the analysis, it is necessary to settle on one control model so that the results can be compared consistently. Of the three control models, the Polynomial and Stepwise were the most true to my assumptions about the Draft. Of those two, I chose to focus on the Stepwise model, which fits the data slightly better and is strictly decreasing. On the issue of three and four year rookie contracts, I decided to use the fourth year salary of players taken from 1995-1997, rather than the fifth year salary. As discussed above, the differences between the two approaches were very small and I decided it was more consistent to always take the year after the rookie contract window. Finally, the easiest decision was to run interval regressions rather than OLS. The interval regression approach is precisely designed to appropriately incorporate heterogeneous censored salary data in the statistical analysis. So, to review: The Draft-order bias control model features:

- Inflation adjusted salaries for 514 NBA players drafted in the period 1995-2003
- Maximum likelihood estimation using an interval regression model, with salaries of league maximum players and NBA dropouts mapped to intervals
- Stepwise controls for Draft pick effects
- Fourth year salaries reflecting players' market values for players drafted from 1995-1997 and fifth year salaries for players drafted from 1998-2003

- Dummies for three weak Draft years: 1997, 2000, and 2002

Using this model I will break down the four biases I detected in the drafting of NBA players.

High School Effect

Starting with the selection of Kevin Garnett in the 1995 Draft, high school players began to pop up on the radars of many NBA teams. Garnett was not the first player ever to be drafted out of high school, Shawn Kemp and Moses Malone are two prior examples, but he began a trend that led to at least one high school player being selected in every subsequent year, until the NBA recently raised its age requirement prior to the 2006 Draft. As Groothius, Hill, and Perri (2005) point out, this trend had a lot to do with the institution of the rookie pay scale, which encouraged teams to pursue riskier players. High school players are certainly riskier prospects since they are younger and have not played against top competition, as college players have. Despite this risk, they also represent an elite level of talent. Players who declare themselves eligible to be drafted out of high school are almost always players who are good enough to be assured of a fairly high selection; otherwise, they would choose to play in college before declaring for the NBA. As a group, high school players who declare for the Draft, create an opportunity to analyze how NBA teams deal with the risk-aversion/option-value dichotomy.

To test for any bias in the drafting of high school players, I inserted a dummy variable for players entering the Draft out of high school.

	1	2	3	4	5	6	7
High School	$944,399 (2.60)			$233,705 (0.29)	$453,840 (0.92)		$224,477 (0.28)
HS Rd. 1		$1,120,648 (2.77)		$888,173 (0.99)		$585,791 (0.95	$363,806 (0.36)
HS			$1,521,27		$1,069,04	$937,05	$934,76

Lottery			4 (2.84)		1 (1.47)	7 (1.15)	3 (1.15)

The first column shows the result of this first regression with high school players returning a large and statistically significant coefficient. The coefficient of $944,399 indicates that, given their position in the Draft, a high school player will generally earn $944,399 more in their next contract than the average player selected at that Draft position. Given the size of this coefficient, I ran further models, in an attempt to isolate where this effect originates. In columns 2-7, there are results for the model when run with additional variables representing high school first round picks, lottery picks, and combinations of the different dummy variables. These models indicate that the bulk of both the high school coefficient value and the statistical significance lie with high school players drafted in the lottery.

Though the actual number of high school players is rather small, the effect is pronounced. Over the period of my data, 1995-2003, there were 22 high school players drafted; 17 of these were taken in the first round, and 11 were lottery picks. Despite the limited sample size, the effect is statistically significant, and it remains significant when the other biases I will be analyzing are added to the model.

	1	2	3	4	5
HS Lottery	$1,521,274 (2.84)	$1,681,296 (3.13)	$1,436,276 (2.69)	$1,507,133 (2.82)	$1,571,251 (2.95)
Center Lot		-$1,098,424 (3.06)			-$936,200 (2.59)
Black			$438,553 (2.14)		$401,490 (1.84)
Black Rd 2			-$606,688 (2.43)		-$547,852 (2.20)
Foreign				-$465,656 (-1.54)	-$390,714 (1.24)
For 95-01				$827,194 (2.10)	$814,509 (2.09)

There is some movement when different variables are added to the model, but the direction of this movement is intuitive and not surprising. When the 'Center Lottery' variable is added, the high school coefficient and z-value both increase. Since the coefficient signs are opposite, the two effects complement each other. This relationship is enhanced by the fact that some of the less successful high school lottery picks were centers. In the next column, when the black player effect is included, you see the opposite movement from the high school lottery variable. In this case, the signs of the two effects are the same; furthermore, all of the high school players in my data were black. This means that the two variables pick up on some of the same things when they are run separately; when run together, their coefficients are somewhat diminished. Regardless, in the final model, which includes all of the effects that I analyzed, the high school lottery coefficient is $1,571,251 and significant at the 99% level with a z-value of 2.95. Clearly, this is a large coefficient, but it can be difficult to decipher exactly what it means. By applying the coefficient to the original control model, we can get a better sense of the strength of this effect.

When we add the high school coefficient to the expected value of a Draft pick based on the control model we can see how the coefficient translates into actual Draft picks. I will refer to the table below to illustrate this analysis.

	Case I		Case II		Case III		Case IV	
Pick	HS	Exp. Value	HS	Exp. Value	HS	Exp. Value	HS	Exp. Value
1	0	$5,546,700	0	$5,546,700	0	$5,546,700	0	$5,546,700
2	0	$5,529,027	0	$5,529,027	0	$5,529,027	0	$5,529,027
3	0	$5,511,354	0	$5,511,354	0	$5,511,354	0	$5,511,354
4	0	$5,493,681	0	$5,493,681	0	$5,493,681	0	$5,493,681
5	0	$5,476,008	0	$5,476,008	0	$5,476,008	1	$7,047,259
6	0	$4,748,610	0	$4,748,610	0	$4,748,610	0	$4,748,610
7	0	$4,730,937	0	$4,730,937	0	$4,730,937	0	$4,730,937
8	0	$3,324,635	0	$3,324,635	1	$4,895,886	0	$3,324,635
9	0	$3,306,962	0	$3,306,962	0	$3,306,962	0	$3,306,962
10	0	$3,289,289	0	$3,289,289	0	$3,289,289	0	$3,289,289
11	0	$2,240,776	0	$2,240,776	0	$2,240,776	0	$2,240,776
12	0	$2,223,103	0	$2,223,103	0	$2,223,103	0	$2,223,103
13	0	$2,205,430	1	$3,776,681	0	$2,205,430	0	$2,205,430

Case I shows the expected value of the top 13 Draft picks based on the Stepwise control model. The values are strictly decreasing, as they always are for this model. If, on the other hand, we assume that one of these players is being drafted out of high school, then we would add the high school coefficient to that expected value. Case II shows the maximum impact that the high school lottery effect can have on a Draft pick. Based on this model, a high school player who is taken 13th is drafted, on average, five spots too low. We can say this because the player's expected value of $3,776,681 is larger than that of the five players drafted directly above him. The team selecting eighth would have been best served by taking him, but the high school lottery bias caused him to fall to the 13th pick. Case III shows the position in the lottery where the high school effect has the least impact. In this case, the expected value of the high school player, $4,895,886, indicates that he should have been taken with the eighth pick instead of the 10th. This movement of two positions represents the smallest possible Draft pick effect that corresponds with the high school lottery coefficient. For the 11 high school players selected in the lottery, the average Draft position, rounded to the nearest whole number, was fifth. A high school player taken at this average value is shown in Case IV. In this scenario the model indicates that the high school player has the highest expected value of any player available. The implication here is that he should have been selected first; hence, he was picked four spots too low. In this way, I have obtained a maximum, minimum, and average amount of Draft pick bias to correspond with the high school lottery coefficient. I will use a similar approach in the other sections of this paper.

Since the Stepwise model moves very unevenly, I also analyzed the results based on the smoother Polynomial model. These results were similar, with the range showing that high school lottery players would be taken four to six picks too low, with the average player taken four spots too low. Four selections may not seem like a lot, but with such a high Draft selection, it can be the difference between a decent player and a franchise player. Among the high school lottery players in my data are great players like Kevin Garnett (1995, #5), Kobe Bryant (1996, #13), Tracy McGrady (1997, #9), and Amare Stoudemire (2003, #9). Missing out on players of this caliber can be a franchise altering event. It is also worth noting that the high school players themselves lose income when they get drafted lower, and end up lower on the rookie pay scale. Recent events have made this issue no longer relevant since the NBA has now instituted an age requirement that prevents players from declaring for the Draft out of high school. My statistical analysis suggests that this NBA rule may, in fact, benefit the players by forcing them to play a year of college basketball, eliminating this high school lottery bias.

These models present strong evidence that risk aversion is a major part of the decision making process of NBA teams. This apparent discrimination against high school players can only be rationally explained by the willingness of teams to pay a premium for the relative security of a college player rather than gambling on a less proven high school player. The fact that this effect is concentrated in the lottery seems to further verify this hypothesis. Unless a pick has been traded, teams drafting in the lottery are teams that have missed the playoffs in the previous season and are in need of improvement. One would assume that a team in this position is less willing to assume risk, given their great need, while a team that had more success in the previous season can afford to take a risk, since they have less immediate need. It seems to follow, logically, that the bias against high school players is strongest among teams drafting in the lottery.

Center Effect

The position of center is the most exclusionary in all of basketball, given the inherent necessity of height that comes with it. Dominant big men have been a major part of the NBA from the very beginning, with players like Kareem Abdul-Jabbar, Wilt Chamberlain, Patrick Ewing, Hakeem Olajuwon, Shaquille O'Neal and Tim Duncan at the center of so many great teams. The conventional wisdom for a long time was that any dominant team needed to have an elite big man. Michael Jordan's Bulls dynasty destroyed this assumption. However, since then, Shaquille O'Neal and Tim Duncan have accounted for a large percentage of the NBA championship titles. The perceived need for a big man is reflected in the Draft. When players like Ewing, Olajuwon, Shaq, Duncan, and recently Greg Oden are available, they are always selected right away. Unfortunately, for every dominant center there are many more that are unable to take their game to the next level. It seems as though every Draft has a center prospect that showed some potential but was a flop in the NBA. In order to test for Draft order bias towards centers, I created dummy variables similar to those in the high school model and added them to my Stepwise control model yielding the following results.

	1	2	3	4	5	6	7
Center	-$145,625 (0.86)			-$70,735 (0.31)	$73,668 (0.39)		-$49,185 (0.22)
Center Rd 1		- $233,461 (0.94)		-$164,770 (0.50)		$330,618 (1.04)	$376,854 (0.98)
Center Lot			-$992,782 (2.75)		-$1,062,644 (2.64)	-$1,301,097 (2.78)	-$1,297,570 (2.77)

The initial results were not at all convincing, but as I segmented the center population I found that there was a pronounced bias in favor of centers selected with a lottery pick. Lottery picks represent a group more than twice as large as the group of high school lottery picks. Of the 119 centers in the data set, 53 were first round picks and 24 were selected in the lottery. I found the result particularly interesting, given the degree to which the effect is concentrated in the lottery. The latter models even show that the effect for centers taken later in the first round (picks 14-29), while not particularly significant, has the opposite sign of the lottery effect. Clearly, there must be a reason, specific to teams drafting in the lottery, which causes centers to be overvalued to the point that when they get a new contract, they earn an average of $992,782 less than other players drafted in the same position.

When other variables are included in the model, the effect remains very strong. While the coefficient and z-value are negatively affected by adding some of the variables, the final model still shows a strong coefficient, with a z-value that is significant at the 95% level.

	1	2	3	4	5
Center Lot	-$992,782 (2.75)	-$1,098,424 (3.06)	-$875,216 (2.41)	-$950,537 (2.63)	-$936,200 (2.59)
HS Lottery		$1,681,296 (3.13)			$1,571,251 (2.95)
Black			$404,970 (1.95)		$401,490 (1.84)
Black Rd 2			-$570,629 (2.27)		-$547,852 (2.20)
Foreign				-$432,953 (1.42)	-$390,714 (1.24)
For 95-01				$780,027 (1.97)	$814,509 (2.09)

The final model shows that centers taken in the lottery earn $936,200 less in their next contract than the average player taken at their Draft position. You may recall from earlier graphs that the Stepwise model predicted a center coefficient of smaller magnitude than the other two control models. This is likely due to smaller salary projections for the top few picks, meaning that the other models may give slightly better coefficient estimates for this particular effect. The coefficients from the other two models are both of magnitude greater than -$1,100,000. These coefficients suggest a statistically significant effect on where centers are drafted.

The next piece of analysis is determining how the coefficient we find translates into actual Draft order. When looking at lottery picks, the Stepwise model gives a range that the center lottery effect could influence Draft position from as little as zero positions to a maximum of being taken seven picks too high. The average center in the lottery was taken with the seventh pick. My Stepwise model suggests that a center drafted seventh would end up being drafted in the correct position, but this result is misleading. Since the seventh pick in this model is right on the edge of a large step down, the center effect is insufficient to reduce the player's value enough so that he falls to the next echelon. In this case, we would get a more representative value by consulting the Polynomial control model. With this model, the center-lottery effect has a minimum impact of a player going four picks too high, and a maximum effect of that player going nine picks too high. At the average Draft position of seven, the effect is that the player was taken five spots too high, so he should have gone 12th. As with the high school effect, this proves to be meaningful when the stakes are as high as they are in the NBA Draft. In the 1998 Draft, overvalued centers Michael Olowokandi (#1) and Robert Traylor (#6) were selected before Dirk Nowitzki (#9) and Paul Pierce (#10). A poor selection in the lottery can be very costly in the long run.

The allure of having a dominant big man is likely a contributing factor in this overvaluing of centers in the NBA Draft. A team that can obtain a franchise center has a great chance of eventually winning a championship; the problem is that such centers are hard to come by. A team drafting in the middle of the lottery may be tempted to Draft a pretty good center and hope that he might be the next great big man; unfortunately, this is never really the case. Teams are so eager for a franchise center, if one is available, he will be drafted first with almost no exception. Despite this, teams continue to take big men later in the lottery hoping they have found a bargain. In reality, if the player was good enough, he would have been taken earlier. This seems like the most likely explanation for this unusual Draft bias. Another possibility is that the NBA game has become faster and more athletic, making centers unnecessary and ineffective. While there is some evidence for this, there are still dominant centers. This rationale would also not explain why the effect only exists in the lottery. Another explanation, which meshes with the hypothesis from the high school effect, is that drafting centers in the lottery is a form of risk aversion. Since height is a known quantity, while talent can be more difficult to judge. Taking a big man may feel like a safer pick. This theory has some appeal, but ultimately it is incomplete. If that explanation were true then we would see a similar coefficient for centers outside the lottery, or at least one with the same sign.

In analyzing these theories it is important to remember that this model assumes that players are paid their market value. The model picks up differences between where players are drafted and what they eventually sign for, hence any bias is specific to the drafting process. For this reason, I find the explanation regarding franchise centers to be the most plausible. While it is a reasonable explanation, it still does not mean the trend is reasonable. It is extremely rare for a franchise center to be found anywhere other than the first overall pick, and for teams to continue to think that they will be successful by overvaluing centers, seems irrational.

Race Effect

As I detailed in my introduction, there is strong evidence from the past showing salary discrimination against black players in the NBA. Current evidence for this is limited, but at the very least the paper by Kanazawa & Funk (2001) indicates that NBA fans do have a bias towards white players. One might expect this bias to show itself in the hiring process in one way or another.

My analysis cannot address racially based pay discrimination because one of the fundamental assumptions I make is that each player is paid his fair market value. If white players are in fact paid more due to their greater marketable production, then my model will take that as their real worth. Adding a race variable to the control model will pick up on any irregularities in the way black players are drafted versus others.[14] The approach here is different from those I described in the introduction and it does not indicate, one way or another, whether there is racial bias in compensation.

When I inserted dummy variables for black players into the model, the result was less than overwhelming; however, when I created separate dummies for black players taken in the first round and those taken in the second round, I found significant results. Not only is there a significant coefficient for black players selected in the first round, it has the opposite sign of the coefficient for black second round picks.

	1	2	3	4	5	6	7	8	9
Black	$139,618 (0.89)								$481,148 (2.33)
Black Rd 1		$517,609 (2.57)				$475,996 (2.15)	$630,873 (1.96)	$481,148 (2.33)	
Black Lottery			$461,223 (1.47)			$154,877 (0.45)			
Black Non-Lot				$368,896 (1.71)			-$154,877 (0.45)		
Black Rd 2					-$257,782 (1.34)			-$158,265 (0.81)	-$639,413 (2.54)

The coefficients here are not as large as for the high school and center effects, but this is partially due to the fact that this analysis deals with the late first round and the second round. Players drafted in these positions have less expected earnings, so these coefficients still represent a significant percentage of these players' anticipated income. It is also interesting to note that the coefficients for black lottery players and black non-lottery players are fairly similar, even though the coefficient represents a larger percentage of the non-lottery player's eventual salary. Another

[14] I say "black" players because my distinction is based on general appearance rather than specific racial background. A dark-skinned player from America, Africa, France, or any other area is considered black while non-black would be comprised of any other players including Americans, Europeans, Asians, etc.

difference between the black player effect and the two previous effects is the sample size, which is significantly larger. Of the 514 players in the data, 362 are black with 182 of them taken in the first round and 180 taken in the second round. Of the four biases I discovered, this one affects the largest group.

When testing this effect with my other variables, I chose to use the dummy variables for black players and black players taken in the second round, the combination from column nine in the above table. The purpose of this combination is that the 'Black' dummy variable picks up the effect for black players taken in the first round, while the 'Black Rd 2' dummy picks up the difference between the first round effect and the second round effect. The coefficients and z-values are slightly weaker when other variables are added, but in the end the results remain meaningful and statistically significant.

	1	2	3	4	5
Black	$481,148 (2.33)	$438,553 (2.14)	$404,970 (1.95)	$528,694 (2.42)	$401,490 (1.84)
Black Rd 2	-$639,413 (2.54)	-$606,688 (2.43)	-$570,629 (2.27)	-$661,098 (2.64)	-$547,852 (2.20)
HS Lottery		$1,436,276 (2.69)			$1,571,251 (2.95)
Center Lot			-$875,216 (2.41)		-$936,200 (2.59)
Foreign				-$456,335 (1.43)	-$390,714 (1.24)
For 95-01				$907,513 (2.30)	$814,509 (2.09)

In the full combination model, the 'Black' coefficient of $401,490 is significant at the 90% level, while the 'Black Rd 2' coefficient is significant at the 95% level. These results strongly suggest that black players who are selected in the first round tend to be drafted lower than they should be. Furthermore, it seems that black players taken in the second round tend to be selected higher than they should be, although this evidence is not overwhelming. The least disputable point of all is that there is a difference between the way black players are drafted in the first round, and the way that they are drafted in the second round. This shift may be the most interesting result of all.

Having obtained these coefficients, we can get a sense of how much movement this bias creates in the Draft itself. Since the analysis is based on full rounds, there are wider ranges of possible effects than we saw with high school players and centers, but the effects are still meaningful. The maximum effect for black first round picks is that they are selected seven positions too low. This occurs towards the end of the round where the slope is flatter. The minimum effect is zero, which occurs at the points where a step drops off in the Stepwise function. The average position for black players taken in the first round is 15th, in which case the player would have been taken an average of four spots too low. The Polynomial model yields similar results, with the minimum effect at one, the maximum effect at seven, and the mean effect at two. Both models indicate that the average black first round pick is selected at least two spots too low.

Even though it is smaller, the black second round coefficient translates to a significant number of picks, since the expected earnings slope in the second round is fairly flat. The Stepwise model has a constant slope for the second round, so the effect is the same for any position in that round. Based on the second round coefficient, the Stepwise model indicates that black players taken in the second round are taken eight spots higher than they should be. For the average black second round position, which is 44th, this means that the player should have been selected 52nd. The stakes are certainly not as high in the second round as they are in the first, but this still represents a substantial overvaluing. Since the Polynomial model is not strictly increasing for the entire second round, it gives a much broader range for the potential bias. It shows the bias could be anywhere from one pick to 21 picks, but interestingly it yields the same value of eight in the case of a player selected 44th overall. With evidence suggesting contrasting black effects for the two rounds of the Draft; all that is left is to speculate as to the root of this peculiarity.

Any Draft bias against black players necessarily implies a bias in favor of white players. There is a popular theory that white players tend to be drafted too high because teams want a franchise player who can be easily marketed to a predominantly white audience. This could provide part of the explanation for the first round effect, but it would seem to conflict with the observed second round effect. It seems that any adequate explanation for this phenomenon needs to simultaneously explain the first round trend and the second round trend. Such a theory would have to be based on an analysis of the inherent differences in the two rounds that might make teams act in completely opposite ways. The two major differences that exist between the rounds are guaranteed contracts and the likelihood of success for drafted players. First round picks receive automatic three year contracts, while second round picks have to fight for a roster spot. Based on my results, I conclude that black players are drafted as a high variance group. That is to say, if a black player and a white player have the same expected value, the black player is perceived to have more upside, as well as more downside. Unfortunately, my analysis cannot address the issue of whether or not this is really the case, but the data indicates that teams perceive this to be the case and Draft accordingly. In the first round, teams are likely to be more risk averse since the opportunity cost is higher in the form of a guaranteed contract. In the second round, the teams should be more interested in option value, since they are not obligated to retain the player. This situation puts a premium on high variance players. Non-black players are taken too high in the first round because teams pay a premium to avoid risk. In the second round, black players are taken too high because teams are willing to pay a premium to obtain greater option value. Whether or not it is a valid generalization, my analysis shows that teams act in a manner that assumes a correlation between race and variance of performance.

Foreign Player Effect

As was the case with high school players, the popularity of drafting foreign players rose significantly during the time of my analysis. The table, below, shows the slow increase in the number of foreign players drafted each year, culminating with large increases in 2002 and 2003.

	1995	1996	1997	1998	1999	2000	2001	2002	2003	Sum
Foreign Players	3	5	6	5	5	9	7	13	19	72

Foreign players selected in the Draft have had mixed results. While some have gone on to be stars, a number of others have been unsuccessful, with some never even joining the team that drafted them. With such an unpredictable group of incoming players, it seems that there is room to tease out some trends governing the way that these players were evaluated and selected.

To my surprise, my initial analysis of foreign players proved to be fruitless. By simply inserting a dummy variable for foreign players into my model, I achieved a coefficient that is almost noteworthy in its insignificance. On the surface, there appears to be a complete lack of evidence for any bias regarding the drafting of foreign players. A deeper analysis, however, shows that this is not the case. Far more foreign players were taken in 2002 and 2003 than in the previous years of my data. This shift provides a useful partition within the larger category of foreign players. When grouped by year, the data provides a much more compelling result.

	1	2	3	4
Foreign	-$17,216 (0.08)			-$495,048 (1.62)
For 95-01			$347,322 (1.30)	$842,370 (2.12)
For 02-03			-$495,048 (1.62)	
For 95		-$113,285 (0.12)		
For 96		$640,290 (0.88)		
For 97		-$319,958 (0.46)		
For 98		-$130,511 (0.18)		
For 99		$543,994 (0.76)		
For 00		$590,977 (1.02)		
For 01		$814,790 (1.34)		
For 02		-$550,697 (1.07)		
For 03		-$462,944 (1.22)		

The breakdown, by year, shows a general scattering of results that grow increasingly positive in 1999, 2000, and 2001. The trend reverses in 2002, when we see a strong negative coefficient that continues into 2003. When grouped in this way, a grouping that is also suggested by the number drafted per year, the data points to a clear trend. From 1995-2001, foreign players in the Draft generally outperformed their Draft position. Then in 2002 and 2003, when more foreign players were selected, they began to underperform relative to expectations. In an approach similar to my handling of the black player Draft bias, I used the 'Foreign' dummy variable to pick up the foreign effect for 2002-2003, and the 'For 95-01' dummy to pick up the difference between the 2002-2003 effect and the 1995-2001 effect.

In a model with the other variables I have discussed, the foreign variables are somewhat diminished, but they do not lose all statistical significance. Neither the 1995-2001 nor the 2002-2003 effect maintains significance, even at the 90% level, but the variable picking up the difference between the two effects is significant at the 95% level in the final combination model.

	1	2	3	4	5
Foreign	-$495,048 (1.62)	-$465,656 (-1.54)	-$432,953 (1.42)	-$456,335 (1.43)	-$390,714 (1.24)
For 95-01	$842,370 (2.12)	$827,194 (2.10)	$780,027 (1.97)	$907,513 (2.30)	$814,509 (2.09)
HS Lottery		$1,507,133 (2.82)			$1,571,251 (2.95)
Center Lot			-$950,537 (2.63)		-$936,200 (2.59)
Black				$528,694 (2.42)	$401,490 (1.84)
Black Rd 2				-$661,098 (2.64)	-$547,852 (2.20)

The result here can hardly be considered a surprise, given the significant increase in the number of players drafted from the first period to the second. It is highly improbable that the amount of international talent increased that abruptly, so it stands to reason that demand for foreign players rose without a similar increase in supply. The result was that teams paid too high a price for these players, who then underperformed relative to these inflated expectations.

Of the four effects I analyzed, this is the most difficult to convert from a coefficient into Draft picks, because the effect is not confined to the lottery or even to just one round. The Stepwise model shows the Foreign 95-01 effect as having a minimum effect of zero at the beginning of the first round, and a maximum effect of 23 positions for players taken in the latter part of the second round. At the average Draft position for these players, which is 32nd, the corresponding effect on Draft position is that the player was taken ten spots too low. The Polynomial model gives a possible bias range from one to 24, and, at the mean value, the bias is also ten. Looking at the 2002-2003 Drafts, we see the same bias range in the Stepwise model due to its constant slope late in the Draft. The bias ranges from zero to 23, but in this case the players are being taken too high. At the mean selection value of 33, players are taken 23 spots too high. The Polynomial model yields similar results, in this case with a range from zero to 25 and a bias of 22 picks at the average selection point in the Draft. The difference in the amount of Draft pick bias that corresponds with the similar coefficients of the two groups, is due to the fact that the slope is much steeper moving up the curve than it is moving down, meaning a coefficient of the same magnitude will create a larger pick effect if it is negative than it would if it was positive. Since there is just one coefficient for both rounds, this analysis is not as insightful when dealing with foreign players as it was for other groups; but, it is clear that the coefficient is meaningful in the sense that it translates to actual drafting error.

It is interesting to compare the way that teams perceive foreign players to their perception of high school players. Based on the bias against high school players, I concluded that teams are risk averse in the Draft, especially in the lottery. One would think that this risk aversion would also apply to foreigners, who have not played college basketball and, hence, cannot be evaluated as easily as their American counterparts. Initially, this appeared to be the case as foreign players were drafted lower than they should have been; a trend that fits neatly into the risk aversion theory. What is interesting is that teams eventually changed their behavior, going so far as to overcompensate and make the opposite error. This change in behavior did not occur in the drafting of high school players, which begs the obvious question of why this is the case. One theory is that the smaller sample size, 22 high school players to 72 foreigners, did not give ample opportunity for teams to change their behavior. Another possibility is that the allure of attracting international fans helped to overcome the risk-averse tendencies of many teams. Perhaps early random successes with foreign players caused teams to overestimate their ability to predict foreign player performance in the NBA. There is also the simple fact that more foreign players were taken later in the Draft, an area where we suspect teams are more willing to accept risk than in the lottery. Either way, the risk associated with foreign players was perceived as either less significant or more tolerable than the risk inherent in high school prospects. This led to the great shift in ideology towards foreign players, from undervaluing them in the period from 1995-2001 to overvaluing them in 2002 and 2003.

Conclusion

My analysis of the 1995-2003 NBA Drafts yielded a variety of results. By creating control models that predict a player's eventual value based on their Draft position, I was able to detect bias related to four distinct groups of players in the Draft. The first of these groups is players who declared for the Draft directly out of high school. These players consistently outperformed their Draft position, especially those drafted in the lottery, and they appear to have been the victims of a bias driven by risk aversion among NBA teams. A second group subject to bias was centers who were selected with lottery picks, although the bias towards them worked in their favor. The root of this bias is somewhat unclear, but it is likely the result of teams deluding themselves in their overeager pursuit of a franchise center. The third group that I found to be subject to bias were black players who consistently out-performed their Draft position in the first round, but under-performed relative to Draft position in the second round. This fits the model of a group that is perceived to be high variance such that risk-averse teams might pass on them in the first round, but prefer them in the second round when their greater perceived option value improves their Draft position. The final group that I found to be subject to bias was foreign players, who over-performed relative to their Draft position from 1995-2001, and then subsequently under-performed as more of them were drafted in 2002 and 2003. Though each effect is derived from the same Draft order bias model, every one of these trends represents a unique issue in the NBA hiring market that points to different biases and tendencies.

Appendix A: Distribution of Players by Draft Year

HS = High School
C = Center
B = Black
F = Foreign

First Round:

	1995	1996	1997	1998	1999	2000	2001	2002	2003
1	B	B	C,B	C,B	B	C,B	HS,C,B	C,F	HS,B
2	B	B		B	B	B	HS,B	B	C,F
3	B	B	B		B	HS,B	F		B
4	B	B	B	B	B	B	HS,C,B	B	B
5	HS,B	B	C,B	B	HS,B		B	F	B
6	C	B	B	C,B		B	B	B	C
7	B	B	B		B	C,B	B	B,F	
8	B	B	C,B	B	B	B	HS,C,B	B	B
9	B	B	HS,B	F	B	C,B	B	HS,B	B
10	B	C,B	B	B	B	B	B	B	B
11	B	C	B	B		B	B	B	B,F
12		C		C	C	C,B	C,F	B	
13	B	HS,B	B	C,B	B	B	B	B	B
14	B	F	B	B	B	B		B	
15			C,B		C,F	C	C,B	F	B
16	B	B	B		B	F		F	B
17		HS,C,B	B	C,F		B		B	C,F
18	B	B	C,F	F	B	B	C,B		B
19	B	B	C		B	C,B	B	B	F
20	B	C,F	C	B	B	B	C,B	B	B
21	B	B	B	B	C	B	B	B	B,F
22	C	B	B	C,B	B	B	B		F
23	B	C,F	B	B	B	HS,B	B	B	HS,B
24	C	B	B	B	F	C,F	F	C,F	B
25	B	F	B	HS,B	B	C,F	B	B	F
26	B	B	B		B	C,B	C,B	B	HS,B
27	B		B	C,F	B	F	B	B	HS,C,B
28	C	C,B,F	B	B		B	B,F		B,F
29	B	C	No Pick	C,B	HS,B		No Pick	No Pick	B

Second Round:

	1995	1996	1997	1998	1999	2000	2001	2002	2003
30	B	B	C	B	B	F	B	B	F
31	C,F		B	B	B		B	B	
32	B		B	HS,B	B	B	B	C	
33	B	B	B	C,B		C	B	B	B
34		B	F	B	C	B	B	B	C,B,F
35	B	B	B	C,F	B,C	B		B	F
36	B		B	C,B	C,F	C,B,F	B	F	B
37	B	B	B,C	C	B,C	B	B	C,F	
38	C,B	C,B	B	B	B		C,F	B	
39	B		B	B	C,B	B	B	B	C,F
40	B	B	B	HS,B	F		B	F	B,F
41		B	C,B	B	C,B	B	B	C,F	B
42	B	B	C,B	B	B	C,B,F	B	B	C,F
43	B	B	B	C,B	B	B	C	C	B
44	C,B	B	B	C,B	C,B		B	B	C,B,F
45	B	C	B	B		C,B	B	B	
46	B	B	B	B		B	C,B	B	F
47	B	B	B	B	C	F	HS,B	C,B	B
48	B		B		B	B	F	B	HS,C,B
49	B	C,B	F	B	C,B		C,B	F	B
50	B	C,B	B,F	C	C,B	B	C,B		B,F
51	F	B		B	B	F	B	F	
52		C	B	B	B	C,B	B	B	C,F
53	C		B	B	B	C	B	B	B
54	C,F	B	C	B	B	B	B	B	C,F
55	C	B	C,B		B	B	B	F	
56	B	B	C,F	B	C	B	F	F	B
57	B			B	F	B	C,B	B	C,F
58	B	C,B	C,F	B	B	B	B	B	C,F

Summary by Category:

	HS	HS Rd 1	HS-Lot	Foreign	Black	Black Rd 1	Black Rd 2	Center	Center Rd 1	Center-Lot
1995	1	1	1	3	44	22	22	10	4	1
1996	2	1	1	5	40	20	20	14	8	3
1997	1	1	1	6	44	23	21	15	7	3
1998	3	1	0	5	41	18	23	15	8	4
1999	2	2	1	5	41	21	20	14	3	1
2000	2	2	1	9	38	19	19	15	9	4
2001	5	4	4	7	46	22	24	14	8	4
2002	1	1	1	13	35	17	18	9	2	2
2003	5	4	1	19	33	20	13	13	4	2
Sum	22	17	11	72	362	182	180	119	53	24

Appendix B: NBA Rookie Scale Contracts, 2003 Draft

Pick	1st Year	2nd Year	3rd Year	4th Opt	5th Qualifying	4th Opt	5th Qualifying
				At 100% Rookie Scale Salaries		%Increases	
1	$3,349,100	$3,600,300	$3,851,500	$4,856,742	$6,313,764	26.10%	30.00%
2	$2,996,500	$3,221,200	$3,446,000	$4,348,852	$5,675,252	26.20%	30.50%
3	$2,691,000	$2,892,800	$3,094,700	$3,911,701	$5,132,151	26.40%	31.20%
4	$2,426,100	$2,608,100	$2,790,000	$3,529,350	$4,655,213	26.50%	31.90%
5	$2,197,000	$2,361,800	$2,526,600	$3,201,202	$4,244,794	26.70%	32.60%
6	$1,995,500	$2,145,200	$2,294,800	$2,909,806	$3,881,682	26.80%	33.40%
7	$1,821,600	$1,958,200	$2,094,900	$2,660,523	$3,567,761	27.00%	34.10%
8	$1,668,900	$1,794,000	$1,919,200	$2,441,222	$3,290,768	27.20%	34.80%
9	$1,534,000	$1,649,100	$1,764,100	$2,247,463	$3,045,313	27.40%	35.50%
10	$1,457,300	$1,566,600	$1,675,900	$2,136,773	$2,910,284	27.50%	36.20%
11	$1,384,400	$1,488,300	$1,592,100	$2,112,717	$2,892,309	32.70%	36.90%
12	$1,315,200	$1,413,900	$1,512,500	$2,084,225	$2,867,894	37.80%	37.60%
13	$1,249,500	$1,343,200	$1,436,900	$2,053,330	$2,839,756	42.90%	38.30%
14	$1,187,000	$1,276,000	$1,365,000	$2,021,565	$2,811,997	48.10%	39.10%
15	$1,127,600	$1,212,200	$1,296,800	$1,987,994	$2,779,216	53.30%	39.80%
16	$1,071,200	$1,151,600	$1,231,900	$1,889,735	$2,655,077	53.40%	40.50%
17	$1,017,700	$1,094,000	$1,170,300	$1,797,581	$2,538,184	53.60%	41.20%
18	$966,800	$1,039,300	$1,111,800	$1,709,948	$2,426,417	53.80%	41.90%
19	$923,300	$992,500	$1,061,800	$1,635,172	$2,331,755	54.00%	42.60%
20	$886,400	$952,800	$1,019,300	$1,571,761	$2,252,333	54.20%	43.30%
21	$850,900	$914,700	$978,500	$1,558,751	$2,246,159	59.30%	44.10%
22	$816,900	$878,100	$939,400	$1,545,313	$2,237,613	64.50%	44.80%
23	$784,200	$843,000	$901,800	$1,530,355	$2,226,666	69.70%	45.50%
24	$752,800	$809,300	$865,800	$1,514,284	$2,213,884	74.90%	46.20%
25	$722,700	$776,900	$831,100	$1,496,811	$2,198,816	80.10%	46.90%
26	$698,800	$751,200	$803,600	$1,448,891	$2,138,563	80.30%	47.60%
27	$678,600	$729,500	$780,400	$1,407,842	$2,087,829	80.40%	48.30%
28	$674,400	$725,000	$775,500	$1,399,778	$2,085,668	80.50%	49.00%
29	$669,500	$719,700	$769,900	$1,389,670	$2,084,504	80.50%	50.00%

Pick #1 receives $3,349,100 in his first NBA year, $3,600,300 in his second NBA year, and so forth. Figures reflect 100% of rookie scale contract salaries. NBA teams have some salary discretion in years 1-3 (salaries can range from 80% to 100% of the rookie scale contract salaries). The percent increases with the exercise of the team's options in years four and five must be at the stated rates.

Source: cbafaq.com

Appendix C: Interval Regression Model

The interval regression model is estimated using Stata's **intreg** command, which is specifically designed to handle interval censored dependent variables. Intreg accommodates censoring that may vary from observation to observation.

This is a maximum likelihood estimation procedure, with observations of the dependent variable either uncensored point data or censored interval data. Let $j \in J$ index the observations, and let $P \subset J$ be the set of point observations and $I \subset J$ be the set of interval observations. For each $j \in I$, then, the interval used in the estimation is $\left[y_{j1}, y_{j2} \right]$, where $y_{j1} = -\infty$ for left-censored data and $y_{j2} = \infty$ for right-censored data.

NBA players at the league maximum salary are right-censored; NBA dropouts fall in the salary interval ranging from 0 to the player-specific league minimum salary.

The underlying model is:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon},\ where\ \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

The log likelihood function with censoring is:

$$\ln L = -\frac{1}{2}\sum_{j \in P}\left\{\left(\frac{y_j - x\beta}{\sigma}\right)^2 + \log\left(2\pi\sigma^2\right)\right\} + \sum_{j \in I}\log\left\{\Phi\left(\frac{y_{j2} - x\beta}{\sigma}\right) - \Phi\left(\frac{y_{j1} - x\beta}{\sigma}\right)\right\}$$

where $\Phi(.)$ is the standard cumulative normal distribution function.

For details see Amemiya (1973).

References

Amemiya, Takeshi. 1973. "Regression analysis when the dependent variable is truncated normal". *Econometrica.* Vol. 41 No. 6, pp. 997–1016.

Becker, Gary. *The Economics of Discrimination*. Second Edition. Chicago: University of Chicago Press, 1971.

Bender, Patricia. "Patricia's Various Basketball Stuff," www.eskimo.com/~pbender/index.html.

Coon, Larry. cbafaq.com

DatabaseBasketball.com

Eschker, Erick, Stephen J. Perez, and Mark V. Siegler. 2004. "The NBA and the Influx of International Basketball Players." *Applied Economics*. Vol. 36, pp. 1009-1020.

ESPN.com

Groothius, Peter A., James Richard Hill, and Timothy Perri. 2007. "Early Entry in the NBA Draft: The Influence of Unraveling, Human Capital, and Option Value." *Journal of Sports Economics*, Vol. 8, No. 3 (June) pp. 223-243.

Hendricks, Wallace, Lawrence DeBrock, and Roger Koenker. 2003. "Uncertainty, Hiring and Subsequent Performance: The NFL Draft." *Journal of Labor Economics*, Vol. 21, No. 4 pp. 857-886.

Kahn, Lawrence M. 1991. "Discrimination in Professional Sports: A Survey of the Literature." *Industrial and Labor Relations Review*, Vol. 44, No. 3 (April), pp. 395-418.

Kahn, Lawrence M. 2000. "The Sports Business as a Labor Market Laboratory", *Journal of Economic Perspectives*, Vol. 14, No. 3 pp. 75-94.

Kahn, Lawrence M., and Malav Shah. 2005. "Race, Compensation and Contract Length in the NBA." *Industrial Relations*, Vol. 44, No. 3 (July), pp. 444-462.

Kahn, Lawrence M., and Peter D. Sherer. 1988. "Racial Differences in Professional Basketball Players' Compensation." *Journal of Labor Economics*, Vol. 6, No. 1 (January), pp. 40-61.

Kanazawa, Mark T., and Jonas P. Funk. 2001. "Racial Discrimination in Professional Basketball: Evidence from Nielsen Ratings." *Economic Inquiry*, Vol. 39, No. 4 (October) pp. 599-608.

USA Today Salaries Database. http://www.usatoday.com/sports/basketball/nba/salaries/default.aspx